DOCUMENT RESUME

ED 194 528 TM 800 290

AUTHOR	Schustack, Miriam W.; Sternberg, Robert J.
TITLE	Evaluation of Evidence in Causal Inference. Technical Report No. 21.
INSTITUTION	Yale Univ., New Haven, Conn. Dept. of Psychology.
SPONS AGENCY	Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE	Oct 79
CONTRACT	N0001478C0025
NOTE	82p.
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Attribution Theory; *Critical Thinking; Higher Education; Hypothesis Testing; *Models
IDENTIFIERS	*Causal Inferences; *Evidence; Inference Skills; Inferential Reasoning

ABSTRACT

In three experiments, this study investigated what evidence people use in making inferences about causality in complex and uncertain situations. Given evidence consisting of multiple observations of some outcome, with each observation including information about the presence or absence of that outcome and of some of its possible causes, subjects estimated the strength of the causal relationship between the outcome and a predetermined possibly-causal event. Over problems and over experiments, the nature and strength of evidence supporting the causal role of the hypothesized cause varied along many dimensions. Using regression-modeling, a set of five evidence types were identified that together gave a good account of subjects' judgments. Four of the independent variables in this model directly concern the relation between the hypothesized cause and the outcome (confirmation by Joint Presence and by Joint Absence of target and outcome, and disconfirmation by violation of sufficiency and of necessity of the target for the outcome), and the fifth represents the goodness of alternative causes as explanations for the outcome. Over the experiments, involving four groups of subjects and five sets of problems, this single linear model accounted for 84 to 90 percent of the variance in each problem-set. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED194528

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Evaluation of Evidence in Causal Inference

Miriam W. Schustack and Robert J. Sternberg

Department of Psychology
Yale University
New Haven, Connecticut 06520

TM800290

Technical Report No. 21
October, 1979

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

This research was sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N0001478C0025, Contract Authority Identification Number NR 150-412.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 21	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) "Evaluation of Evidence in Causal Inference"		5. TYPE OF REPORT & PERIOD COVERED Periodic Technical Report (1 Jul 79 - 30 Sep 79) 6. PERFORMING ORG. REPORT NUMBER Research Report No. 7-79
7. AUTHOR(s) Miriam W. Schustack and Robert J. Sternberg		8. CONTRACT OR GRANT NUMBER(s) N0001478C0025
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology Yale University New Haven, Connecticut 06520		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61153N; RR 042-04; RR 042-04-01; NR 150-412
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE 1 Oct 79 13. NUMBER OF PAGES 63
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Portions of this paper were presented at the annual meeting of the American Psychological Association, New York, September 1-5, 1979.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Causal inference, induction, base rate

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In three experiments, we investigated what evidence people use in making inferences about causality in complex and uncertain situations. Given evidence consisting of multiple observations of some outcome, with each observation including information about the presence or absence of that outcome and of some of its possible causes, subjects estimated the strength of the causal relationship between the outcome and a predetermined possibly-causal event. Over problems and over experiments, the nature and strength of evidence

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

supporting the causal role of the hypothesized cause varied along many dimensions. Using regression-modeling, we found a set of five evidence types that together gave a good account of subjects' judgments. Four of the independent variables in this model directly concern the relation between the hypothesized cause and the outcome (confirmation by Joint Presence and by Joint Absence of target and outcome, and disconfirmation by violation of sufficiency and of necessity of the target for the outcome), and the fifth represents the goodness of alternative causes as explanations for the outcome. Over the experiments, involving four groups of subjects and five sets of problems, this single linear model accounted for 84 to 90% of the variance in each problem-set.

Evaluation of Evidence in Causal Inference

Miriam W. Schustack
Carnegie-Mellon University

Robert J. Sternberg
Yale University

Short Title: Causal Inference

Send Proofs to:
Miriam W. Schustack
Department of Psychology
Carnegie-Mellon University
Pittsburgh, Pa. 15213

Abstract

In three experiments, we investigated what evidence people use in making inferences about causality in complex and uncertain situations. Given evidence consisting of multiple observations of some outcome, with each observation including information about the presence or absence of that outcome and of some of its possible causes, subjects estimated the strength of the causal relationship between the outcome and a predetermined possibly-causal event. Over problems and over experiments, the nature and strength of evidence supporting the causal role of the hypothesized cause varied along many dimensions. Using regression-modeling, we found a set of five evidence types that together gave a good account of subjects' judgments. Four of the independent variables in this model directly concern the relation between the hypothesized cause and the outcome (confirmation by Joint Presence and by Joint Absence of target and outcome, and disconfirmation by violation of sufficiency and of necessity of the target for the outcome), and the fifth represents the goodness of alternative causes as explanations for the outcome. Over the experiments, involving four groups of subjects and five sets of problems, this single linear model accounted for 84 to 90% of the variance in each problem-set.

Evaluation of Evidence in Causal Inference

In science and in daily life, we are frequently faced with the problem of determining whether, or to what extent, a particular action or event is responsible for some observed outcome. Will "deeper" semantic processing produce better memory? Will Head Start programs improve school achievement? Will telling jokes ensure the success of lectures? In such situations, many aspects of the data could be relevant to evaluating a particular cause, for example: how often the hypothesized cause and the outcome co-occur, the presence and strength of alternative possible causes, the relative frequency of occurrence of the outcome, the *a priori* likelihood that this cause would have an effect on this outcome, the number of observations available, the proportion of missing data, etc. What types of evidence do people perceive as supporting (or weakening) a hypothesized cause, and what are the relative weightings among these different types of evidence?

In the domain of philosophy, J. S. Mill (1843) proposed a set of heuristics for determining the necessary and sufficient causes of an outcome event, given information about the presence and absence of that outcome and of its possible causes for some number of situations. Mill's "methods of experimental inquiry," derived from Hume's "rules by which to judge of causes and effects" (1739/1888), prescribe what Mill believed to be optimal strategies for inductive reasoning, given that necessity and sufficiency are the only kinds of causal relations to be

considered, and given observations of every possible cause in every observed situation. For example, the simplest of Mill's methods, the "direct method of agreements," allows the determination of a cause necessary for the presence of some outcome, given (a) a fixed set of possible causes, (b) two or more situations where the outcome is present, and (c) information about the presence and absence of each possible cause for each situation. After assembling these observations, one looks for a cause that is present in every situation (or absent in every situation); if there is such a cause, that cause (or its absence) could be necessary for the outcome.

Other philosophers, such as Carnap (1962) and Scriven (1976), have also suggested approaches to evaluating causal relations, providing guidelines for determining what evidence is necessary to confirm a hypothesized cause. In the social psychological literature, but in much the same vein as Mill, Kelley (1967) has suggested that attributions of causality can be made using an analysis-of-variance model. Using this approach, one assembles multiple situations involving different combinations of possible causes (from a fixed set) that are present, and records the outcome for each. One then looks for the "effects" of the individual causes (as for variables in an ANOVA), and thus determines which cause is responsible for the outcome. This scheme as well requires complete information, and assumes that there is no limitation on the amount of information that can be handled. Both of the above-mentioned proposals (Mill's and Kelley's) are quite reasonable methods of analyzing causal evidence, given complete information and unlimited

processing capacity.

Clearly, though, these assumptions of optimal conditions do not appropriately characterize either the problem or the problem-solver in naturalistic causal inference-making. What do people without training in logic actually do when confronted with this kind of problem? Constrained by their "bounded rationality" (Simon, 1957), subjects cannot be expected to make perfectly rational attributions of causality (Fischhoff, 1976). In causal reasoning, people are subject to illogical and statistically unsound biases that cannot be explained by any simple notions of processing limitations (cf. Tversky & Kahneman, 1977); these biases affect what information the subjects attend to, and what they do with the information they do choose to consider. For example, subjects judging contingency between two events overvalue confirming instances, and thus overestimate the relationship between those events (Jenkins & Ward, 1965; Smedslund, 1963). In making generalizations, subjects are much more influenced by instances that confirm their hypotheses than by disconfirming instances (Gollob, Rossman, & Abelson, 1973; Lord, Ross, & Lepper, in press). Their assessment of data is influenced by pre-existing schemata and associations (Chapman, 1967; Chapman & Chapman, 1967, 1969; Kelley, 1972; Crocker & Taylor, Note 1). When confronted with an outcome having multiple possible causes, subjects assess causal relationships employing a simplifying principle of "minimum causation" (Shaklee & Fischhoff, Note 2), rather than a more complex principle of discounting (Kelley, 1972): In these multi-causal situations, subjects interpret the known presence of one possible cause as sufficiently ex-

plaining the outcome, and they greatly decrease their estimated probability of the presence of other, supposedly independent, possible causes.

One question that these findings raise is whether all the above-mentioned effects will occur (and how the ones that do occur will combine), given complex problems in which subjects themselves must select the information to consider, with instructions and stimuli that highlight neither particular strategies nor particular portions of the evidence. When the subjects must choose what aspects of the problem to consider in judging a causal relation, what sorts of data do they consider, and how strongly do each of these data types influence perceived causality?

The experiments reported below attempt to specify how subjects judge the probability that a hypothesized cause would produce a particular outcome event when the subjects are given incomplete information about complex problems that vary simultaneously on many dimensions. Since subjects had to choose what aspects of the evidence to consider, and how to weigh and combine these evidence types, our research may be more faithful to the conditions of causal judgments in the real world than is much of the earlier work. The complexity of the problems, the minimally constrained task, and the use of multiple-regression techniques for analysis make this research particularly informative. Taken together, these features allow us to assess (a) the relative importance of inherently-correlated evidence types; (b) the roles played by biases and limitations such as those mentioned earlier; and (c) the weights as-

signed to different kinds of evidence (within single experiments).

The paradigm used in these experiments is one of evaluating a preselected causal hypothesis--the task that we, as scientists, perform in evaluating experimental data, and the task that also characterizes a large portion of our causal reasoning in everyday life. The technique to be used to analyze the data from these experiments is one of testing various models of subject performance via multiple regression, and comparing these models as accounts of the data. The nature of our experimental task and of our analytical technique enabled us to encourage subjects to use their (presumably complex) real-world strategies in solving the problems. In this research, then, we can find evidence about some questions whose previous investigation was limited to situations where the variable of interest was necessarily highly salient to subjects.

This research can be seen as having at least one goal in common with some of the social psychological research on attribution: Both seek to understand how people determine the causes of events. The two types of research differ, however, in what comparisons are of interest. The attribution literature is focused on questions of the extent to which and the conditions under which subjects attribute outcomes to different kinds of causes (e.g., internal vs external, self vs others, situations vs dispositions)—distinctions that are orchogonal to the issues that most interest us in this program of research. We are primarily interested in how subjects use patterns of evidence to determine the strength of a causal relation, and how subjects integrate the prior

strength of a causal hypothesis with their evaluation of the evidence presented. In our attempts to tap subjects' use of their prior knowledge about real-world domains (in Experiment 3), we contrast causes that differ in their prior strengths, but that are still of a uniform nature with respect to dimensions that are of importance to social psychology. The flavor of the attribution research is thus quite different from that of the experiments reported here, but the two approaches are complementary; perhaps at some point in the future, they will converge. [For a review of empirical work on attribution, see Nisbett and Ross (in press). See also Jones, Kanouse, Kelley, Nisbett, Valins, and Weiner (1972) for many important original papers in the field.]

In a series of three experiments, subjects evaluated some causal hypothesis in each of many problems. A sample of these problems appears

---

Insert Table 1 about here

---

in Table 1. In each problem, subjects were presented with the hypothesis that a particular event was responsible for some outcome, and were asked to use a given body of evidence to estimate the probability that the hypothesized causal event, present in isolation. would produce that outcome. Each line of a problem depicts one "situation"; in each situation, the result (to the right of the arrow) is the occurrence or non-occurrence of the outcome event. The hypothesized cause to be evaluated (the "target") is always specified (offset below the problem). Within a situation, each possible cause (the target and all alterna-

tives) is either present, absent, or not observed.

Within each experiment, problems varied in terms of the number of situations in which the target was present, the number in which it was absent, and the number in which it was not observed; over each of these situation types, the outcome (occurrence vs. non-occurrence of the outcome event) also varied. Additionally, the problems varied in terms of the number of situations observed, the number of possible causes observed per situation, the number of different alternative causes presented in the problem, and the strength of the evidence supporting these alternatives.

Other features varied only between experiments. In Experiment 1, there were never any situations in a problem that disconfirmed the target's causal role (that is, there were no violations of necessity or sufficiency of the target to the outcome); in Experiments 2 and 3, these disconfirmatory situations were permitted. In Experiments 1 and 2, the problems consisted of only abstract patterns of events, represented by letters of the alphabet; in Experiment 3, some sets of problems involved concrete events from real-world domains.

The organization of the remainder of this paper will be as follows: First, there will be a general description of each experiment. Then, there will be a description of the method employed for all the experiments, and of the special features of each experiment. Following this, the modeling of all the data will be discussed. This organization will allow for the most coherent presentation of the results of the regression modeling, since evaluation of the models for each experiment should

be done considering not only the results of that experiment alone, but the data from all the experiments as a whole.

In the first experiment, the hypothesis to be evaluated was never directly disconfirmed by the evidence--if the target (hypothesized) cause was observed, it and the outcome were either both present or both absent. These simpler problems thus involved no direct violations of necessity or sufficiency relations between the target cause and the outcome. The permissible evidence types were restricted in this way because of the possibility that the presence of any problems with directly disconfirmatory evidence might have a major effect on subjects' strategies. The limited nature of the problems in this experiment provide an opportunity to observe a somewhat simpler situation, which can later be compared to the more complex situation that subjects face in the second and third experiments. In Experiment 1, subjects had to evaluate and combine information about varying levels of confirmation, but the only forms of evidence that reflected negatively on the target hypothesis were the absence of confirmation and the strength of alternative hypotheses. In the later experiments, subjects had to integrate these forms of evidence with observations that were directly disconfirmatory of the target hypothesis.

The second experiment differed from the first in two major features. First, it allowed the presence of evidence types directly disconfirming the target hypothesis (violations of sufficiency and necessity). While these evidence types were not present in every problem, the general effect of their presence in the overall set can be as-

sessed by comparing the models and parameter estimates from this experiment with those of Experiment 1.

The second major difference from the first experiment was that, in addition to our continued focus on group data, here we were also interested in looking at the performance of individual subjects. Although the reliability of responses between subjects in Experiment 1 was very high (as we will discuss later), we were not fully convinced that there were no interesting individual differences in performance on this type of task. In the second experiment, then, we attempted to increase our chances of finding differences between subjects, if there were any to be found. In order to get more data, and more useful data, about the performance of individual subjects, we made two modifications to the basic experiment. First, there were additional replications of each basic problem (in this case, four observations of each basic problem instead of two). These additional replications increased the stability of the independent variable for individual-subject modeling (a subject's mean response to a basic item), and also provided more data on within-subject consistency over replications.

As another way of getting more information about individual subjects, we added a set of tests of inductive reasoning ability. These tests provide a means of relating individual performance in our causal-inference task to performance in other domains of inductive reasoning. The ability tests were of the type used in intelligence-test batteries, and thus differed in one important respect from the causal-evaluation task: they were scorable in terms of whether or not each response was

the correct answer (or at least, the generally accepted "best" answer) to the problem. Inclusion of this rather different type of task allows us to address the issue of whether the kinds of inductive abilities required for successful performance in these "objective" tests are related to performance in the less-structured evaluation task.

In the third experiment, as in the second, all possible relations between the target and the outcome were permitted. In this experiment, we included sets of problems that consisted of abstract symbols (letters of the alphabet) representing the possible causes and the outcome, just as in the first two experiments. Here, though, we also included sets of problems where each abstract problem was realized as a set of situations involving concrete events.

The third experiment thus served two purposes. First, with its abstract problems, it provided an independent replication of the results of the second experiment, with new subjects and new items. Given that the primary form of data analysis for these experiments was multiple regression, it was important to guard against the possibility that the models selected were largely capitalizing on chance variation in the data (Cohen & Cohen, 1975).

The second (and more important) contribution of the third experiment is that it provided a measure of how subjects combine their abstract evaluation strategies with their prior knowledge in problems involving concrete domains. One of the important motivations for introducing the concrete content was to examine the effects of the a priori probability of a hypothesized cause, when subjects have to evaluate the

causal role played by that hypothesized cause in some particular set of data. In order to look at how prior probabilities were combined with the abstract structure of the evidence, we needed to know what these prior probabilities were. Thus, an independent group of subjects provided estimates of the base rate or a priori probability for each concrete potential cause.

In the main task, where subjects evaluated entire problems, the factor of concreteness of the items to be judged was varied between subjects for two reasons: There was a possibility that the presence of meaningful content in any of the items that a subject judged might alter that subject's stratgies for the abstract items, thus weakening the ability of the abstract portion of this experiment to serve as a replication of Experiment 2. On the other hand, it was also possible that the absence of meaningful content in any of the problems might alter strategies for the concrete items, and thus prevent us from finding differences in the treatment of concrete vs. abstract items that would otherwise exist.

METHOD

All Experiments

Subjects

Subjects were members of the Yale University community, who participated for pay ($2.50 per hour), course credit in Introductory Psychology, or a combination of pay and credit. Forty-three subjects participated in Experiment 1, 58 in Experiment 2, and 123 in Experiment 3. Of the subjects in Experiment 3, 40 were given abstract-content problems,

62 were given concrete-content problems, and 21 estimated *a priori* probabilities of concrete target hypotheses.

Materials

Materials for all experiments were causal inference problems requiring an evaluation of the causal role of a single possibly causal event on the basis of a pattern of evidence. In each experiment, there was a large set of problems, and each was presented with replication within subjects, using isomorphic forms of each problem set. For abstract-content problems, the isomorph sets were simple alphabetic transformations of one another (e.g., every R in Form 1 became an M in Form 2); an independent, random alphabetic transformation was used to produce each isomorph set. For concrete-content problems, the isomorph sets involved two different real-world domains.

Each problem consisted of two parts: the presentation of evidence, and the identification of a possible cause to be evaluated. The subjects' task was to "use the evidence in each problem to determine the likelihood that a particular one of the possible causes, in isolation, leads to the outcome." A sample problem (of one of the types used in the experiments) appears below:

A B → E

-A C → -E

B C → -E

C -A → E

A → E

Evidence. The evidence consisted of descriptions of a number of

situations (where each situation occupied a single line of the problem); each situation consisted of a number of observed (possibly causal) events and an outcome. For example, in the problem above, A, B, and C are the possibly causal events, and E is the outcome event.

Over the set of problems in each experiment, the number of situations described in a single problem varied between two and five, and the number of possibly causal events that were observed per situation also varied between two and five, independently of the number of situations. Within a single problem, each situation had the same number of observed possibly-causal events.

Within any situation, each possible cause (e.g., A in the sample problem above) was in one of three states: observed to be present (e.g., A in the first line), observed to be absent (e.g., -A, read "not A," in the second line), or not observed (e.g., the third line, where no mention is made of A as either present or absent). For all situations in a single problem, the outcome event (e.g., E in the sample problem) was always the occurrence (e.g., E in the first line) or the non-occurrence (e.g., -E in the second line) of the same outcome event. Exhaustive combination of the three conditions of each possibly causal event (positive, negative, not-observed) with the two outcome conditions (positive and negative) yields six different relationships that can exist between each possibly-causal event and the outcome event for any one situation (line) in the problem. In terms of the implications of these six relationships for the causal role of the observed, possibly causal events, these relationships are:

1. Confirmation by positive outcome: The event is observed to be present and the outcome occurs. This relation between the event and the outcome is evidence for the sufficiency of the event for the outcome.

2. Disconfirmation by negative outcome: The event is observed to be present but the outcome fails to occur. This relation between the event and the outcome is evidence against the sufficiency of the event for the outcome.

3. Disconfirmation by positive outcome: The event is observed to be absent but the outcome occurs. This relation between the event and the outcome is evidence against the necessity of the event for the outcome.

4. Confirmation by negative outcome: The event is observed to be absent and the outcome fails to occur. This relation between the event and the outcome is evidence for the necessity of the event for the outcome.

5. Uninformative positive outcome: The event is not observed and the outcome occurs.

6. Uninformative negative outcome: The event is not observed and the outcome fails to occur.

Target Hypothesis. In each problem, the subjects were to evaluate the causal role of one possibly causal event, the "target hypothesis." Below the evidence in each problem we designated the possibly-causal event that was the hypothesized cause of the outcome event (e.g., A → E in the example problem). The task was to estimate the likelihood (probability converted to a 0 to 100 scale) that the presence of the

target event, in isolation, would lead to the occurrence of the outcome event.

Over the set of basic problems in each experiment, the number of occurrences (and the proportion of occurrences) of each of the permissible evidence types for the target event was varied, as were the number (and proportion) of each of the permissible evidence types for the alternative causes.

Procedure

Subjects were tested in groups comprising between 2 and 15 subjects. Subjects were given printed instructions, which were read aloud by the experimenter while subjects read them silently. The instructions pointed out the complex and uncertain nature of the causal evaluation task, stressing that subjective estimates rather than correct calculations were required. Subjects were told to consider each problem independently, and to assume that there was no relation between the events in one problem and the events in any other problem. They were told not to go back to consult any previous answers. The instructions highlighted the difference between definite absence of a possible cause and lack of information about it. Also mentioned were the possibility of unobserved events affecting the outcome, and the limited nature of the available evidence (in terms of number of possible causes observed in each situation, as well as in number of situations). The similarity between the experimental situation and the real world (in terms of the limitations of the evidence and the complexity of the task) was repeatedly pointed out.

The experiments were conducted as paper-and-pencil tasks. Problems were presented in booklets, and subjects recorded their responses on prepared answer sheets. Subjects were permitted to mark up the problem booklets or to make notations if they so desired. The task was self-paced, with the experimenter telling subjects when one-quarter, one-half, and three-quarters of the estimated time had elapsed. There was some variability in finishing times, with subjects in Experiment 1 and the abstract condition of Experiment 3 spending about one hour, subjects in Experiment 2 and the concrete condition of Experiment 3 about 90 minutes, and subjects in the Base Rate condition of Experiment 3 about half an hour. All subjects were allowed adequate time to complete all problems.

Design

Within each of the five subject groups (Experiment 1; Experiment 2; and the three groups of Experiment 3, to be called 3-Abstract, 3-Concrete, and 3-Base Rate) every subject received all sets of problem isomorphs, with the order of administration of forms counterbalanced over subjects on the basis of order of arrival at the experimental session. The order of the problems within each form was randomized independently for each subject. The subject completed all problems of one isomorph set before another set was begun.

Experiment 1

In Experiment 1, the evidence involving the target event was restricted to only those evidence types that did not directly disconfirm a causal relationship between the hypothesized cause and the outcome (that

is, to evidence types 1, 4, 5, and 6). There were no instances of evidence type 2 or 3, since these violate sufficiency and necessity of the target for the outcome. For all the other possibly causal events in a problem, though (that is, for the alternative causes or hypotheses), all evidence types were permissible (types 1 through 6).

Experiment 1 involved two isomorphic sets of 60 basic problems.

Experiment 2

The causal inference problems in Experiment 2 differed from those of Experiment 1 in that they allowed situations exemplifying all six of the logically-possible relations between the target hypothesis and the outcome. Thus, not only were there confirmatory and uninformative evidence types involving the target cause (types 1 and 4, and types 5 and 6, respectively), but there were disconfirming evidence types (types 2 and 3) as well. Instructions for this experiment were altered to include a warning that interactions between possible causes might produce an apparent relation between a possible cause and the outcome that differed from the true relation.

This experiment involved four replications (isomorphic sets) of 50 problems. The isomorph sets were counterbalanced over subjects in a Latin Square design. Following the last problem set, subjects answered a questionnaire about their strategies.

Reasoning Tests

Experiment 2 included an additional set of materials: three brief paper-and-pencil inductive reasoning tests. These were a geometric-series completion task, a letter-series completion task, and a geometric

analogies task. In the geometric-series completion task, items required a forced choice for items of the form A:B:C:D:? involving spatial relations between geometric forms. The letter series task was of the same type, but involved alphabetic progressions. In the analogies task, items were of the form A:B::C:?, with 5 response-alternatives, and involved spatial and geometric relations. The inductive reasoning ability tests, which were timed (10, 5, and 5 minutes for the geometric series, letter series, and geometric analogies, respectively), were administered before the self-paced causal inference task. The three reasoning tests were given in random order for each group.

Experiment 3

The problems in Experiment 3 involved the same full set of evidence types used in Experiment 2: All six evidence types were permissible for the target cause and for all alternative possible causes.

3-Abstract

Although the abstract problems in Experiment 3 (those using letters of the alphabet for the events) were constructed independently of those for Experiment 2, they met the same set of restrictions. Here, though, there were 60 problems, and two isomorphic sets. The instructions from Experiment 2 were used for these subjects as well.

3-Concrete

The problems in the concrete-content condition of Experiment 3 were formally identical to the abstract-content problems of that experiment. There were two concrete-content domains: epidemics of diseases and declines in stock values. For each content domain, a set of 26 possibly-

causal events was assembled, where each event was a plausible cause of the type of outcomes found in that domain (that is, disease epidemics or stock-value declines). Examples of these events can be found in the concrete-content problems in Table 1. These events were arbitrarily assigned to letters of the alphabet, and the evidence in each problem was constructed by substituting a concrete event for the corresponding letter in the abstract version of each basic problem.

A particular concrete outcome event was also required for each problem: the particular disease involved in the epidemics, or the particular industry in which stock values were dropping. For the disease epidemics, we invented a set of 60 fictional diseases with names that gave no information about the nature of the disease, such as Dempes Syndrome and Manley-Lincoln Disease; these diseases were then randomly assigned to the 60 problems that concerned epidemics. For the stock crashes, we constructed a set of 60 types of corporations covering a wide range of industries, such as pharmaceutical companies, publishers, resort operators, and fruit packers. These were assigned to the 60 problems with the constraint that the kinds of possible causes mentioned in a problem had to be applicable to the type of business assigned to that problem: For example, certain of the causes we used are sensible only for manufacturing-type industries, others only for service-type industries.

The problems in each content domain were set up in a format appropriate to that domain. (See Table 1 for sample problems.) For the disease epidemics, each line of the abstract problem was translated into

the events or conditions present (or absent) in one city, with the outcome being an epidemic (or no epidemic) of the disease assigned to that problem. For the stock crashes, each line of the abstract problem was translated into the events or conditions present (or absent) in one company of the industry assigned to that problem, with the outcome a drastic drop (or no drastic drop) in the value of that company's stock.

For the epidemic domain, the judgment requested was the probability that in some other city, an epidemic of the assigned disease would occur if the hypothesized cause were present. For the stock domain, the judgment requested was the probability that in some other company of the assigned industry, there would be a drastic drop in the value of stock if the hypothesized cause were present.

Each subject in the concrete-content group solved both the set of 60 epidemic problems and the set of 60 stock-crash problems, with order of administration counterbalanced over subjects. The instructions for this condition varied only slightly from those in the abstract condition. They prepared subjects to judge problems from "two different domains," and provided a sample problem from a concrete content domain, although a different one from those the subjects would judge--the ratification of a constitutional amendment. The instructions also encouraged subjects to use their prior knowledge about the causes involved, directing subjects to "keep in mind all the information available to you--information both in the problem and already known to you."

3-Base Rate

The base-rate problems were the same as the concrete-content prob-

lems of Experiment 3; except that each problem provided no evidence, and gave only the conclusion to be judged. Subjects in this condition saw questions of the form, "What is the probability that, in some city, there would be an epidemic of Phipp's Disorder if diseased cattle were slaughtered and sold?", and "What is the probability that, for some coffee company, stock values would drop drastically if there was political upheaval at the company's foreign location?".

Here, the instructions told subjects that they would be asked to judge the likelihood of some outcome's occurring under certain circumstances. Subjects were given no other information about how to determine the answer, except that they were to try to use the full range of the rating scale.

RESULTS

Basic Statistics

The mean response for each data set, over subjects and over items,

---

Insert Table 2 about here

---

appears in Table 2. These mean responses appear, on preliminary inspection, to be quite similar. Since Experiments 1, 2, and 3 involved different sets of basic items, we did not perform any comparisons of mean response across experiments. For the three data sets of Experiment 3, however, two comparisons are reasonable: (a) a comparison between responses to abstract-content problems and responses to concrete-content problems (between subjects), and (b) a comparison between responses to epidemic-domain problems and responses to stock-domain problems (within

subjects).

For the first comparison, we calculated the average response for each subject (collapsing over forms and items) and compared the means for the abstract-content group with the means for the concrete-content group. The difference between groups, with means of 34.99 and 36.20 respectively, was not significant, t (100) = .62. This absence of difference suggests that subjects in the two groups were using similar decision rules in making their judgments; additional, more fine-grained analyses will explore the similarities.

For the second comparison (epidemics versus stocks), the subjects' mean responses to epidemic problems were compared with mean responses to stock problems, by a matched-sample t-test over subjects. The mean responses to epidemic problems were significantly lower than the mean responses to stock problems (both given in Table 2), t (61) = 2.87, p < .01. Given that the estimates from the base-rate subjects were also much lower for the epidemic problems (33.96, SD = 14.71) than for the stock problems (44.65, SD = 13.33), the difference for subjects given the full problems can be attributed to differences in the base-rates. Further analyses will clarify the importance of differing base-rates.

Reliability

One of the striking aspects of the data we collected was their high reliability. Evaluation of the adequacy of our regression models requires estimates of how much reliable variance there was in each data set. We measured reliability in a number of ways--between forms (isomorphic sets), between replications (orders of presentation), between

subjects, and within subjects. By all of these measures (presented below for each data set), our data appeared to have very high internal consistency.

Experiment 1

Using mean responses over subjects, the correlation between forms of the problems, adjusted by the Spearman-Brown formula, was .98; the corresponding correlation between replications (here, the first- vs the second-presented set) was .97. Consistency of responses over subjects was assessed using Cronbach's coefficient alpha (Cronbach, 1951). For this experiment, alpha was equal to .96. Since these measures of reliability all showed the responses to be little influenced by form, order, or individual subject, it was appropriate to do regression modeling of this experiment on the basis of a mean response to each of the basic problems, averaged over these other three variables.

Experiment 2

In this experiment, pairwise correlations among the four forms and among the four replications were also very high (adjusted mean r both .99). Cronbach's coefficient alpha showed the reliability over subjects (collapsing over forms) to be .99.

In this experiment, where we modeled individual as well as group data, we looked at the reliability of responses from individual subjects. We assessed within-subject reliability by averaging each subject's responses to Forms 1 and 2, and averaging his responses to Forms 3 and 4. For each subject, we correlated these two 50-item sets, and adjusted the correlations by the Spearman-Brown formula. The mean

of this reliability coefficient, over subjects, was .83 (SD = .11).

Experiment 3

3-Abstract. In this data set, the adjusted correlations between forms and between replications were both .98, and coefficient alpha was .99.

3-Concrete. Reliability of responses to these problems was assessed in three ways: collapsed over content domains, within each content domain, and between content domains. Collapsing over domains, coefficient alpha was .99. Reliability within each domain was assessed by correlating the responses, collapsed over subjects, to a domain when it was the first-presented set with responses to that domain when it was the second-presented set (a between-subjects comparison). This correlation, adjusted by the Spearman-Brown formula, was .98 for disease epidemics and .97 for stock crashes. Although this reliability is quite high, evidence from the between-domains comparisons suggests that the apparent reliability within domain is due to the reliability of responses to a given basic problem (that is, to the abstract aspect of the evidence) rather than to subjects' consistent use of the base-rate probabilities: The correlation between concrete domains was very high (r = .97). As another measure of this same aspect of the reliability, the correlations between the responses to the abstract problems and responses to each concrete domain were also very high: r = .98 and .97 for epidemics and stocks, respectively.

3-Base Rate. The base-rate estimates themselves were quite reliable, as assessed by the procedures used above for the concrete-content

problems. For the base-rate problems, the correlation between orders of administration (adjusted by Spearman-Brown) was .93 for the epidemics and .91 for the stocks. For this base-rate condition, there was, of course, no relation between a problem in one domain and the corresponding problem in the other domain, since responses here were just ratings of the target hypotheses, which were randomly assigned to the basic problems. The correlations above demonstrate that there was great consensus among subjects about the relative strengths of the causal relations between the various causes and their outcomes.

Models

Proposed Model of Group Data

The model we propose as the best account for subject performance over all three experiments combined is one consisting of five independent variables (plus the regression constant, which represents what the level of response would be in the absence of evidence). These variables, taken together, extract a large portion of the potentially relevant information in each problem; they are sensitive to the observations in a problem that support and oppose the target hypothesis, as well as taking into account the evidence in favor of competing possible causes. These variables represent:

1. confirmation of the target hypothesis by joint presence of target and outcome (evidence type 1 described earlier),

2. disconfirmation of the target hypothesis by violation of sufficiency of the target for the outcome (evidence type 2),

3. disconfirmation of the target hypothesis by violation of neces-

sity of the target to the outcome (evidence type 3),

4. confirmation of the target hypothesis by joint absence of target and outcome (evidence type 4),

5. disconfirmation of the target hypothesis by the strength of alternative hypotheses.

For each problem, levels of the independent variables for the first four evidence types were determined by simply counting the number of situations in that problem that fell into each of the four categories. Levels of the fifth variable were determined by computing an index of causal relatedness (to the outcome) for each alternative variable in a problem, and then taking the average of the two highest scores on that index. The index consisted of the sum of the number of observations confirming the causal role of some alternative hypothesis minus the number of observations disconfirming the causal role of that hypothesis. Confirmations and disconfirmations were based on the relation between the alternative and the outcome, with joint occurrences and joint absences considered confirmation, and mismatches in presence and (known) absence considered disconfirmation. By this procedure, a variable was computed for each problem that represented the average strength of the two alternative causes that were the best competing explanations for the outcome.

The adequacy of the proposed model can be evaluated in many ways: by its fit to the data (in terms of variance accounted for and unaccounted for), by the plausibility and comparability over experiments of the model's parameter estimates, and by the model's superiority over plausi-

ble alternative models. These criteria will be disussed in the sections immediately below. Additional criteria, such as the model's relation to data from other research, and more global features of the model, will be covered in the "General Discussion" section.

Goodness-of-fit. As the initial portions of Table 3 show, this

---

Insert Table 3 about here

---

model provides a good account for the responses in all five sets of data (Experiments 1, 2, 3-Abstract, 3-Stocks, and 3-Epidemics). In terms of the proportion of the variance accounted for, this model provided a uniformly good fit to data from three completely independent experiments, involving both different subjects and different problems.

Limitations of the proposed group model. We should point out that this model is being espoused not as the "true" model for these data, but only as the best of the alternatives on the basis of our criteria. Comparing the R-squared for each data set with its reliability (given earlier), there is clearly a sizable gap between the proportions of variance accounted for (.84 to .90) and the reliabilities (all over .96 by different measures).

The failure of the proposed model to capture all the systematic variance in the data is also demonstrated by an analysis of residuals. We separated out two non-overlapping subsets of responses for each of the five data sets. These subsets consisted of responses, collapsed over all relevant subjects, either to half the isomorph sets of the problems (Experiments 1, 2, and 3-Abstract), or to half the replications

of the problems (Experiments 3-Stock and 3-Epidemic). Separate regression analyses were run for each subset, and residuals were recorded on an item-by-item basis (predicted minus observed value). The correlation between the sets of residuals was significant for each of the five data sets, with all coefficients greater than .75, based on 50 or 60 pairs of residuals (all $p < .001$). If the variance not accounted for by the model were not systematic, these residuals would represent only error, which would be expected to show no systematic relationship across subsets of data. Given these large and significant correlations between residuals, we must conclude that there is systematic variance in the responses that our preferred model is failing to capture.

Comparability and plausibility of parameter estimates. One measure of the goodness of the proposed model is the extent to which the same independent variables are significant for each of the data sets. Table 3 gives the regression coefficients for each independent variable; all the independent variables included were significant in every experiment, at $p < .01$ or better, with one exception: The Joint Absence evidence type for Experiment 3-Stocks was significant only at the .05 level. (The two directly disconfirming evidence types did not enter into the model for Experiment 1, of course, where no direct disconfirmations were permitted in the problems.)

The adequacy of the proposed model can also be evaluated by the extent to which a single group of independent variables displays comparable and reasonable parameter estimates for each data set in which it appears. In the case of the current model, we would minimally require

that the two confirmatory evidence types would take positive coefficients, and the three disconfirmatory types, negative coefficients. This expected pattern is confirmed (see Table 3).

Beyond this initial test of reasonableness, close inspection of the coefficients further supports our model. The relationships between the coefficients within any one experiment tended to be repeated for the other experiments as well. For example, the weight given to positive (Joint Presence) confirmations tends to be much larger than the weight given to negative (Joint Absence) confirmations. With the exception of Experiment 2, where the values for Joint Presence and Joint Absence are nearly equivalent (for reasons to be discussed below), there is a sizable bias in favor of Joint Presence as a positive contributor to the probability that the outcome will occur given the presence of the target cause.

Perhaps, given the particular question that subjects were asked in each problem, this lopsided view of confirmatory evidence is not surprising. At an abstract level, either form of confirmation should increase a person's subjective estimate of the contingency between target and outcome. The Joint Presence instances, though, also provide exemplars of the actual relationship to be judged, that of joint presence. Joint Absence may, in a statistical sense, be as informative as Joint Presence with respect to the strength of the causal relation, but joint presence can be seen as more relevant to the prediction of future joint presence.

As mentioned earlier, Experiment 2 fails to follow this pattern of

unequal valuation of confirmations by joint presence versus joint absence. This deviation from the results of the other four data sets can be explained in terms of some unfortunate limitations in the problems for that experiment. The variability of the confirmatory evidence types for Experiment 2 was extremely small (with standard deviations of .70 and .40 for Joint Presence and Joint Absence, respectively) in comparison to the variability of those evidence types in the other experiments (1.31 and 1.16 for Experiment 1; 1.29 and .83 for Experiment 3). In fact, Experiment 2 had the smallest variation between problems for every evidence type except Strength of Alternatives.

This restriction of the variance of the independent variables may have made the parameter estimates somewhat misleading, and may account for the fact that the R-squared of the proposed model was lower for Experiment 2 than for all three data sets of Experiment 3 (.84 vs .90, .88, it .90); and was even lower than for Experiment 1 (.84 vs .85), in which the two disconfirming independent variables were absent. Relative to Experiment 3-Abstract, the more restricted range of the values for the evidence types in Experiment 2 led to less variability in the dependent measure. As shown in Table 2, there was less variability between problems in the mean response over subjects: the standard deviations were 14.96 versus 18.90, significantly different from one another ($F(59,49) = 1.60$, $p < .05$). Thus, the reduced variability in the dependent measure seems to explain the smaller R-squared values and the somewhat deviant parameter estimates in the second experiment.

Additional evidence for the consistency of our proposed model over

all experiments, and for its reasonableness as a psychological model as well, comes from comparing parameter estimates for the two disconfirming evidence types. Table 3 shows that violations of sufficiency ($\underline{A} \rightarrow -\underline{E}$) were taken to be more damaging to the target hypothesis than were violations of necessity ($-\underline{A} \rightarrow \underline{E}$) in all four data sets where these evidence types appeared. Philosophers have argued for centuries about the relative contributions of necessity and sufficiency to causality (see Beauchamp, 1974; Sosa, 1975). In our experiments, however, the task was more specific than that of evaluating "causality"; our subjects were evaluating the probability of the occurrence of the outcome in the presence of the target alone, clearly a judgment more dependent on the sufficiency of the target for the outcome than on its necessity for the outcome. Given this task, it is reasonable for subjects to be more concerned about violation of sufficiency, which is the relation to be judged, than about some other violation of contingency. As was the case with the two types of confirmation, these two types of disconfirmation may both contribute to the assessment of a causal relationship, but one (in this case, violation of sufficiency) may have been perceived as more relevant to the question at hand.

Alternative Models

In this section, we will cover the comparative evaluations of the proposed model, describing some of the other models and variables we tested, and how they fared as accounts of the data. First, we will discuss alternative representations of evidence directly relevant to the hypothesized cause, and then we will discuss alternative representations

of information relevant to the causal role of competing causes.

Evidence directly relevant to the target. One plausible method of evaluating the evidence in our problems is by considering the proportion of target-outcome relations that were of each type. Although the proposed model involves counting the number of occurrences of each evidence type, there are good arguments in favor of measuring their proportion instead. Using proportional independent variables, the response measure becomes less dependent on the number of observations in the problem. We explored some of the possible conceptualizations of "proportional information," each with its own rationale. For each of these alternative models, the "Strength of Alternatives" variable was included for comparability to the proposed model.

1. One proportional approach is to take the number of instances of each evidence type and divide it by the total number of situations (lines) in the problem. This measure is grounded on the notions that (a) the situations presented are a sampling of a larger set, and (b) despite the small numbers, these observed situations are representative of the larger set. The model based on these simple proportions as independent variables did not fare well: The R-squared measures were below those of the proposed model for each data set, as shown in Table

---

Insert Table 4 about here

---

4. The simple proportional model results in an average loss of .05 in variance accounted for relative to our preferred model.

2. Another conception of proportionality involves discarding si-

tuations that are uninformative relative to the target (evidence types 5 and 6), and proportionalizing over the number of informative situations. This model is even poorer than the simple-proportion model above. The R-squared values, given in Table 4, showed that this model was worse in every case, for an average decrement of .12.

3. We also considered the possibility that the evidence types should be seen in relation to the number of observations with the same outcome--for example, that the prevalence of Joint Presence is best measured by the proportion of Joint Presence situations among all situations where the outcome occurs. Subjects did not seem to be using this approach: The R-squared measures for this model were lower for every experiment than those of the preferred model, with an average decrease of .12.

4. The last of this set of models was a combination of the two models immediately above, that is, discarding uninformative situations and then dividing by the number of remaining situations with the same outcome. This approach could not be tested for Experiment 1: Given the absence of evidence types 2 and 3, these variables would represent Joint Presence over Joint Presence, and Joint Absence over Joint Absence, and thus would either have the value of 1, or be undefined (zero over zero). We did test this model for the four data sets of Experiments 2 and 3, and it resulted in R-squared values that were lower for all (mean loss of .17).

5. In another vein entirely, one can conceive of the appropriate value for each evidence type depending on the number of variables (whose

presence or absence is known) that could have been responsible for the outcome in each situation. This would mean, for example, that the appropriate conception of a Joint Presence situation depends on the number of variables observed in that situation: The observation would be more informative with respect to the target if there were, for example, only two possible causes rather than five. Thus, each independent variable in this model was derived by dividing the number of occurrences of each evidence type by the number of variables observed per situation (which was constant within a problem). This model did not do as well as our proposed model for any data set--the R-squared values were, on the average, .08 below those of the preferred model.

All five of the above alternative approaches, then, compare unfavorably with our proposed model. We also tested one more approach, somewhat different from the ones above: a Bayesian estimation of the probability. This approach involved calculating the probability of the outcome given the presence of the target, using Bayes' Theorem. Given that the number of observed situations per problem was at most five, the Bayesian approach may be inappropriate, or too crude to be useful. Perhaps this explains why the independent variable representing the Bayesian estimate was so poor: In models including the Bayesian variable along with the variables of our proposed model, the Bayesian quantity did not even enter the equation. We can conclude, then, that subjects were not using a Bayesian approach in their evaluation, a frequent finding in the judgment literature (cf. Tversky & Kahneman, 1974).

We also tested a number of other approaches in an attempt to see if

there was any nonlinearity in how subjects were affected by the number of occurrences of each evidence type. These included using nonlinear transformations of the number of occurrences of each evidence type (e.g., reciprocals and square roots), separating out the first occurrence of each evidence type from subsequent occurrences (to see if these were unequally weighted), and converting the number of occurrences into a binary variable (no occurrences vs. any occurrences) for each evidence type. All of these approaches resulted in a significant decrement in R-squared, relative to the proposed model. Thus, results for the alternative models just described provided no support for the notion that the variance left unaccounted for by the proposed model reflected nonlinearity in the subjects' consideration of evidence, although, of course, we tested only a few of the possible forms nonlinearity might take.

Evidence relevant to competing causes. Other aspects of the evidence could also be perceived in ways that differ from the proposed model. The proposed model has a single variable representing the strength of competing causes. We tried other alternatives in which individual evidence types relevant to the best competing hypotheses, to the arithemetic average of the best two competing hypotheses, and to a weighted average of these two were added to the models. All effects were too weak to be statistically significant. The summary variables, such as the one in the proposed model, where all the evidence relevant to a competing hypothesis is collapsed into a single index of causal strength, fared much better. We tested three of these variables: (a) the proposed Strength of Alternatives variable, representing the average

strength of the best two competing causes, (b) a variable representing only the single best competing hypothesis, and (c) a variable representing a weighted average (3:1) of the summary indices of the best two competing hypotheses. The patterns of superiority for these were somewhat mixed over data sets. The variable we finally chose as the best for the proposed model was a compromise; we cannot claim that it is obviously superior in all cases.

An alternative way of looking at competing hypotheses is to assume that covariation between target and competitors is informative. That is, if there was a competing cause that covaried to a large extent with the target, the responsibility of the target for the outcome is diminished. We found no evidence that subjects were sensitive to this covariation; numerous indices measuring covariation failed to contribute significantly to our equations.

Another possibility is that subjects considered only those competing causes that were superior or equal to the hypothesized cause in their relation to the outcome. This approach was apparently not used, though: Independent variables representing this approach (in a few different ways) were inferior to the proposed Strength of Alternatives variable (which evaluates alternatives independently of whether or not they are better explanations than the target cause).

Another possible way of taking competing causes into account is by simply counting how many different ones there are in a problem. This approach, however, appeared not to have been used by the subjects--a variable representing this quantity failed to reach significance in all

data sets.

Models of Individual Subject Data

Our modeling of individual data was designed to serve two purposes. First, we wanted some measure of the extent to which the group model was an appropriate representation of the performance of individuals. To accomplish this, we wanted to identify subjects whose responses might have been better described by models other than the preferred group model. Second, we were also interested in relationships between subjects' behavior in evaluating causal evidence and other measures of inductive reasoning ability, as assessed by looking at relations between a subject's individual model and scores on the other inductive reasoning tests.

Fit of the proposed model to individuals. We performed a regression analysis on each individual subject's data, using the same set of independent variables as in the proposed group model. Overall, the group model did a good job of accounting for the data from individual subjects. The mean over subjects in proportion of variance accounted for was .64 (SD = .15), for data that had a mean reliability over subjects of .83. The mean over subjects of the proportion of reliable variance accounted for (R-squared divided by reliability) was .77 (SD = .16).

Our attempts to find subjects who were likely to be using models other than the proposed one were not very successful. There was a strong correlation between R-squared for the proposed model and the reliability of the subject's data, r (58) = .50, p <.001. One result of

this was that, in general, subjects who had poor fits to our proposed model were assured by the low reliability of their data of having poor fits to any model. To circumvent this problem, we separated out the group of subjects with the largest amount of unaccounted for systematic variance, to see if we could find evidence of their uniform and consistent use of some other model. Inspection of this measure (reliability minus R-squared) identified six subjects whose scores clustered above the remainder of the group. We correlated the response sets of these six subjects among themselves. Agreement was quite low: The mean correlation between all possible pairs of these six subjects was .34. (This mean can be compared with the results for the six subjects with the smallest amount of unaccounted-for systematic variance, where the mean correlation was .64.) Thus, the subjects poorly fit by our model appear to have used idiosyncratic or inconsistent decision rules. There was no evidence for their use of some other shared model.

Reasoning tests. Scores on the three reasoning tests were highly correlated (all $r > .52$). A factor analysis on these scores was performed to provide a single index (factor score) of reasoning ability for each subject, to be used in conjunction with that subject's regression modeling. The factor analysis yielded a single factor, with approximately equal weightings on the three test scores. Since this derived reasoning factor correlated almost perfectly with a simple mean of standardized scores on the three tests ($r = .99$), the composite z scores were used for all further analyses involving the reasoning tests.

The correlations with our ability-test data were disappointing.

Beyond the correlations among the test scores and factor scores, there were no significant relations to the estimated parameters of the model, model fit, or reliability.

Modeling Concrete-Content Items

As mentioned earlier, there was little effect of base rates on subjects' responses, beyond the possible effect of higher base rates producing higher responses for the stock problems than for the epidemic problems. It appeared from the between-domain comparisons of Experiment 3-Concrete that subjects were very strongly influenced by the abstract pattern of the evidence. This does not, however, preclude the base-rate probabilities of the hypothesized causes having some effect on responses, although such an effect would surely be secondary. Some simple correlational measures imply that there is a role played by the base-rate probabilities. If there were such an effect, it would appear in a measure of the deviation of each concrete problem from its corresponding abstract problem. Correlating this deviation score for each problem with the base-rate probability for that problem, we obtain an estimate of the extent to which deviation of the concrete from the abstract was due to the base-rate probability of the target hypothesis. Taking a simple difference between concrete and abstract versions of each problem, and then correlating this with the base-rate for that problem, yields significant correlations: overall, r (120) = .41, $p < .001$; for the epidemics, r (60) =.41, $p < .001$; for the stocks, r (60) = .31, $p < .02$.

When a regression was performed with the base rate as an additional predictor variable, along with the variables of the proposed model, base

rate did not add statistically significant prediction to the model. In fact, even if the predictors for the concrete-content item responses were reduced to two--the response to that problem in the abstract form and the base rate--the effect of base rate was still almost nil: For the stocks, the base rate was a non-significant parameter, while for the epidemics, it was significant (at $p < .001$), but added less than .02 to the R-squared. Clearly, the incremental impact of the base rates was very small.

We also calculated a variable to represent a different possible incorporation of the base-rate: a multiplicative one. This variable, though, failed completely, not entering the models significantly for either concrete set.

One possible explanation for this neglect of base rates is that the use of two different domains led subjects to feel that they should disregard them, even though the instructions encouraged them to use their prior information about the causes. This explanation probably does not account for our findings, though. Assuming that this contamination due to multiple domains affected the second-presented set exclusively (or at least, primarily), we would then expect base-rates to be more important for the first-presented set for each subject. We found no larger role of base rates when we modeled responses to each content-type when it was the first-presented set than when it was the second-presented set, nor was the correlation between the base rate and the deviation from abstract responses any higher for the first-presented set than for the second-presented set--it was, in fact, nonsignificantly

lower (.27 vs .37). Thus, subjects were no more sensitive to base rates in the concrete domain they saw first.

This neglect of base rates is not surprising or unprecedented: A large literature provides examples of other situations where base rates have been statistically relevant, but ignored by subjects (e.g., Lyon & Slovic, 1976; Nisbett, Borgida, Crandall, & Reed, 1976). There are situations, though, where subjects have been shown to be sensitive to base rates (e.g., Ajzen, 1977; Carroll & Siegler, 1977). One explanation that has been offered for the pattern of use versus nonuse of base-rate information is that of Tversky and Kahneman (1977): They propose that a causal interpretation of the base rate (as opposed to a diagnostic one) is required for subjects to attend to base rates. That hypothesis appears to predict that our problems would involve the use of base rates, since the relation between the base rate and the judgment being made seems appropriately causal. Perhaps a better explanation for the neglect of base rates in our case is one similar to a proposal of Nisbett & Borgida (1975): When subjects have a great deal of other information from which to determine their responses, they tend to neglect base rates. In our problems, subjects may have felt that the information presented about the role of the hypothesized cause was more nearly complete and more relevant, even perhaps more reliable, than any intuitive estimation of the base rate probability for that cause.

Alternatively, the apparent nonuse of base rates by subjects in our study might have been due to weaknesses in problem-construction: The base rates for our concrete causes may not have differed sufficiently

from one another. Although the different causes were assigned a wide range of base rates by subjects in Experiment 3-Base Rate (from 10.81 to 69.84 on a 0-to-100 scale), the restriction of possible causes to ones that were plausible may have made the true range of the base rates used quite small--too small for effects of base rates to be significant.

Subjects' insensitivity to the manipulation of abstract versus concrete items has some interesting implications for performance with abstract-content problems. One could argue that the abstract-content problems present to the subjects a highly impoverished stimulus. But, subjects did not appear to use the additional information available when the problems were made richer and more complex by using concrete causes. Thus, for the abstract problems, the reliability of responses and the simplicity of the variables accounting for responses is not readily attributable to the poverty of the presented information.

In the same vein, the similarity between abstract and concrete versions of the problems casts doubt on the notion that subjects were doing estimations of correlation rather than of causality when given abstract problems. The presentation of items with concrete content would make causal relations more salient than the abstract items would: The relation between mosquito infestation and disease epidemics, for example, is hardly bidirectional. If subjects were sensitive only to correlation when given abstract problems, one would expect differences between responses to abstract versus concrete items on the basis of the saliency of the causal relation for the concrete-content problems. It might well be that causal analysis is routinely done using the same methods by

which people assess correlation--all we are claiming here is that subjects in our experiments were probably using some version of whatever procedures they normally use for evaluating causality.

Similarly, one could argue that responses to concrete problems reflected a search for higher-order causes to explain both the target cause and the outcome, a strategy that would be feasible only for the concrete-content problems. The absence of a difference between concrete and abstract problems makes it implausible, though, that the possibility of higher-order causes had any significant effect on responses.

GENERAL DISCUSSION

Features of the Proposed Model

Earlier, we discussed the adequacy of the proposed model in terms of its overall goodness of fit, the comparability and plausibility of the model's individual independent variables and their weights, and the model's superiority over plausible competing models. In this section, we will discuss the adequacy of the model in terms of the relation between the proposed model and data from other tasks, and the extent to which the model meets criteria of uniformity, completeness, and simplicity.

Parallels with Other Data

A comparison of our results with the existing literature on subjects' behavior in other hypothesis-testing tasks shows good concordance. When subjects are proposing tests of some hypothesis, there is a strong bias for seeking out and attending to positive confirmations as the most informative type of datum. This occurs both for subject-

generated hypotheses (e.g., Wason, 1960; 1968; Wason & Johnson-Laird, 1972), and experimenter-determined hypotheses (e.g., Snyder & Swann, 1978).

In our experiments, there is a comparable overall focusing on positive confirmation: The Joint Presence evidence type is the single best predictor of response, as shown by three measures. First, Joint Presence is the first variable to enter the stepwise regression for all data sets. Second, it has the highest individual regression coefficient F statistic of all the variables in the proposed model, for four of our five data sets (the exception was Experiment 2, where Joint Presence suffered from limited variability). Third, in the fitting of the proposed group model to individual subject data, Joint Presence was the independent variable that reached significance for the largest number of subjects (90%).

The focus on positive confirmation can be seen as evidence of a more general bias against negativity. For example, the preference of our subjects for Joint Presence over Joint Absence as confirmatory evidence is consistent with the psychological literature on many forms of inductive reasoning, where information in a negative form is undervalued relative to its true importance: Most relevant here is the research on concept identification (e.g., Bruner, Goodnow, and Austin, 1956) and on contingency estimation (e.g., Smedslund, 1963).

Combining portions of the information presented earlier, we can draw two generalizations about how negativity affects subjects' hypothesis evaluation. First, information in a positive form (Joint

Presence vs Joint Absence) tends to be more important to subjects than information in a negative form. Second, information with a positive meaning (confirmation vs disconfirmation) tends to be more important to subjects as well. These preferences can be seen as manifestations of a more general preference for positivity over negativity, found not only in many other forms of inductive reasoning, but also in other cognitive domains (e.g. language comprehension, sentence verification, deductive inference, social inference, and language development).

Uniformity, Completeness, and Simplicity

One positive feature of the proposed set of variables is that they are all basically of one type. The four variables directly relevant to the target hypothesis are computed in a uniform fashion, and the fifth variable can be seen as (more or less) a combination of these same evidence types, in this case from the perspective of the competing (and thus individually less important) causes. Given the range in the manner of computation of the variables tested for alternative models, the proposed model presents variables that are highly consistent with one another in terms of what they measure and how they measure it.

This is not to imply, however, that the variables in our model are the only variables, or the only kinds of variables, that affect causal inference. On the contrary, it seems likely that other kinds of variables account for at least some of the systematic variance left unexplained by the proposed model.

Another dimension to consider is that of completeness of the model. Together, the five variables in the proposed model encode most of the

information relevant to the causal relation between target and outcome, or at least are sensitive, in some limited fashion, to most of that information. We do not claim that these variables cover all possible methods of assessing this causal relation. Nor do we claim that these variables represent precisely and completely the decision rules used by subjects; the proposed model does not account for all systematic variance. We do, however, assert that the overall model is made more plausible by its completeness; our model reflects (at least partially or indirectly) most of the important aspects of the evidence--including those suggested to us by other research and by our subjects in their reports of the strategies they used.

Accepting that subjects have limited capacity to process information in a given problem, a model with few parameters, whose variables are similarly computed and encompassing of the evidence, has more psychological plausibility than one whose variables are both diverse and selective. In addition to these properties, discussed above, another psychologically important dimension is simplicity of the variables, or ease of encoding. One feature of the proposed group of variables is that, relative to many of the alternatives considered, the five variables are quite easy to calculate for any individual problem. In many of the alternatives we considered, the computation of a single variable could require multiple operations of selection, accumulation, and division. In the proposed model, the set of variables is relatively easy to compute: The four direct evidence types (which concern the relation between target and outcome within each observation in a problem)

can be computed by finding the sign of the target variable and of the outcome in each line, and counting the number of occurrences of each type of relation. Computation of the fifth variable, Strength of Alternatives, requires two executions of this procedure, and a combination of the counters.

We would not want to claim, though, that subjects necessarily computed their responses in precisely the same manner as the model does. The strongest claim we make is that, among the variables we considered, the proposed set of independent variables produced the quantity that best captured the subjects' evaluations; the subjects' method and the model's method of arriving at that evaluation may or may not be quite different.

Given that we are modeling only the final responses, we cannot interpret our variables as accounts for processes executed while subjects are working on a problem. On the other hand, a model of response choice that fails to consider processing implications of the variables included is of limited usefulness. While the model we are supporting here was derived only from response-choice data, the model can be evaluated with a more global view. The features discussed in this section (uniformity, completeness, and simplicity) are relevant to evaluating the proposed model as a basis both for response choice and for the processing that underlies response choice. Clearly, any definitive statements about the computations subjects performed, or the duration or sequencing of these processes requires more and different data from that collected in these studies. If, though, our ultimate goal is to understand how people make

judgments of causality from evidence, the response-choice model must be sensitive to processing implications of the proposed variables.

CONCLUSIONS

Our proposed model, with its implications about how subjects evaluate evidence in judging a causal hypothesis, makes certain general claims about people's inductive behavior. Clearly, people are not optimally rational, in the sense of extracting the statistically maximal amount of information from the evidence they are given. They are subject to many biases, as detailed in the discussion of the models: for example, biases against negativity in any form, biases in favor of evidence that more closely matches the superficial form of the statement to be evaluated, biases in favor of the hypothesized cause over other causes. Most of these biases can be accounted for by a mechanism such as focusing--the "irrational" aspects of behavior in this task seem quite explicable in terms of the subjects' focusing on their task, and on the most salient individual components of the evidence (cf. Taylor & Fiske, 1978).

And yet, subjects' behavior was not wholly predictable by bias: Many of the features of our data depict the subject as a rather rational evaluator, taking into account many different aspects of the evidence presented, and using these appropriately. While their evaluation seems to include some deviant weightings of the information in the problems, the information they choose to attend to is not grossly deviant from what seems appropriate. Their focusing behavior may represent an adaptive method of dealing with limitations of memory and processing capaci-

ty. For example, subjects might believe that the strength of a causal relation is better measured by the proportions of each evidence type than by the number of each, but they may be unable (or unwilling) to do the additional computation that sensitivity to proportionality requires. Overall, given their limitations, subjects perform this kind of causal inference task in a reasonable, if imperfect, manner.

Reference Notes

1. Crocker, J., & Taylor, S. E. Theory driven processing and the use of complex evidence. Paper presented at the meeting of the American Psychological Association, Toronto, September 1978.

2. Shaklee, H., & Fischhoff, B. Limited minds and multiple causes: Discounting in multicausal attribution. Unpublished manuscript, University of Iowa, 1977.

References

Ajzen, I. Intuitive theories of events and the effects of base-rate information on prediction. Journal of Personality and Social Psychology, 1977, 35, 303-314.

Beauchamp, T. L. (Ed.). Philosophical problems of causation. Encino, Calif.: Dickenson, 1974.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: Wiley, 1956.

Carnap, R. Logical foundations of probability (2nd ed.). Chicago: University of Chicago Press, 1962.

Carroll, J. S., & Siegler, R. S. Strategies for the use of base-rate information. Organizational Behavior and Human Performance, 1977, 19, 392-402.

Chapman, L. J. Illusory correlation in observational report. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 151-155.

Chapman, L. J., & Chapman, J. P. Genesis of popular but erroneous psychodiagnostic observation. Journal of Abnormal Psychology, 1967, 72, 193-204.

Chapman, L. J., & Chapman, J. P. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. Journal of Abnormal Psychology, 1969, 74, 271-280.

Cohen, J., & Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, N.J.: Erlbaum, 1975.

Cronbach, L. J. Coefficient alpha and the internal structure of tests.

Psychometrika, 1951, 16, 297-334.

Fischhoff, B. Attribution theory and judgment under uncertainty. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), New directions in attribution research. Hillsdale, N.J.: Erlbaum, 1976.

Gollob, H. F., Rossman, B. B., & Abelson, R. P. Social inference as a function of the number instances and consistency of information presented. Journal of Personality and Social Psychology, 1973, 27, 19-33.

Hume, D. Treatise of human nature. (L. A. Selby-Bigge, Ed.) London: Oxford, 1888. (Originally published, 1739).

Jenkins, H. M., & Ward, W. C. Judgment of contingency between responses and outcomes. Psychological Monographs, 1965, 79, (1, Whole No. 594).

Jones, E. E., Kanouse, D. E., Kelley, H. H., Nisbett, R. E., Valins, S., & Weiner, E. (Eds.). Attribution: Perceiving the causes of behavior. Morristown, N. J.: General Learning Press, 1972.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), Nebraska symposium on motivation (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. Attribution in social interaction. Morristown, N. J.: General Learning Press, 1972.

Lord, C. G., Ross, L., & Lepper, M. R. Biased assimilation and attitude polarization: The effects of prior theories. Journal of Personality and Social Psychology, in press.

Lyon, D., & Slovic, P. Dominance of accuracy information and neglect of

base rates in probability estimation. Acta Psychologica, 1976, 40, 287-298.

Mill, J. S. A system of logic, ratiocinative and inductive. London: J. W. Parker, 1843.

Nisbett, R. E., & Borgida, E. Attribution and the psychology of prediction. Journal of Personality and Social Psychology, 1975, 32, 932-943.

Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H. Popular induction: Information is not necessarily informative. In J. S. Carroll & J. W. Payne (Eds.), Cognition and Social Behavior. Hillsdale, N. J.: Erlbaum, 1976.

Nisbett, R. E., & Ross, L. Human inferences: Strategies and shortcomings in social judgment. Englewood Cliffs, N. J.: Prentice Hall, in press.

Scriven, M. Maximizing the power of causal investigations: The modus operandi method. In G. V. Glass (Ed.), Evaluation studies: Review annual (Vol. 1). Beverly Hills: Sage Publications, 1976.

Simon, H. A. Models of man. New York: Wiley, 1957.

Smedslund, J. The concept of correlation in adults. Scandinavian Journal of Psychology, 1963, 4, 165-173.

Snyder, M., & Swann, W. B. Jr. Hypothesis-testing processes in social interaction. Journal of Personality and Social Psychology, 1978, 36, 1202-1212.

Sosa, E. (Ed.) Causation and conditionals. London: Oxford, 1975.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top

of the head phenomena. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 11). New York: Academic Press, 1978.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Tversky, A., & Kahneman, D. Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), Progress in social psychology. Hillsdale, N. J.: Erlbaum, 1977.

Wason, P. C. On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 1960, 12, 129-140.

Wason, P. C. 'On the failure to eliminate hypotheses...'--A second look. In P. C. Wason & P. N. Johnson-Laird (Eds.), Thinking and reasoning. Hammondsworth, England: Penguin, 1968.

Wason, P. C., & Johnson-Laird, P. N. Psychology of reasoning: Structure and content. London: D. T. Batsford, 1972.

Footnote

This research was supported by Contract N0001478C0025 from the Office of Naval Research to Robert Sternberg and by a National Science Foundation Graduate Fellowship to Miriam Schustack. We thank Barbara Conway, Elizabeth Charles, Elizabeth Hartka, and Ed Barksdale for assistance on this project, and Susan Fiske for comments on an earlier version of the manuscript. Portions of this article were presented at the annual meeting of the American Psychological Association, New York City, September 1979. Requests for reprints should be sent to Robert J. Sternberg, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, Connecticut 06520.

Table 1

Sample Problems

for All Experiments

Experiment 1

A	B	-C	→	G		L	T	M	R	Q	→	F
-A	D	C	→	-G		L	T	-M	R	Q	→	-F
B	E	C	→	-G								
C	D	A	→	G								
										M	→	F

A → G

Experiments 2 and 3-Abstract

R	T	H	Q	→	B		P	X	→	Y	
Q	-R	L	T	→	-B		N	X	→	-Y	
-Q	-T	U	H	→	-B		N	P	→	Y	
L	Q	-R	T	→	B						
-T	-H	R	Q	→	-B					P	→ Y

R → B

Table 1 (Continued)

Experiment 3-Concrete (Epidemics)

An epidemiologist noted that, for Marshall-Isaacs disease

In City 1

Annual health inspection of food-service workers was stopped

A new type of hair dye was introduced to the area

No inadequate sterilization practices were used in a local cannery

A new type of pesticide was tried by local vegetable farmers

There was a water main break

An epidemic of the disease was reported.

In City 2

Inadequate sterilization practices were used in a local cannery

Anuual health inspection of food-service workers was not stopped

Lead-based paint was sold without warning labels

There was no water main break

A new type of pesticide was tried by local vegetable farmers

No epidemic of the disease was reported.

What is the probability that, in some other city, there would be an epidemic of Marshall-Isaacs disease if a new type of hair dye was introduced to the area?

Table 1 (Continued)

Experiment 3-Concrete (Stocks)

A market analyst noted that, among cosmetic manufacturers:

In Company 1

The office staff of the company organized and joined a union

The company's major product was under suspicion as a carcinogen

There was a drastic drop in the value of the company's stock.

In Company 2

The office staff of the company did not organize or join a union

The company's major product was under suspicion as a carcinogen

There was a drastic drop in the value of the company's stock.

In Company 3

Illegal campaign contributions were traced to the company's managers

The company's major product was not under suspicion as a carcinogen

There was no drastic drop in the value of the company's stock.

What is the probability that, for some other cosmetic manufacturer, stock values would drop drastically if the company's major product were under suspicion as a carcinogen?

Table 2

Characteristics of Likelihood Judgments

for All Experiments

	Experiment				
	1	2	3-Abstract	3-Stock	3-Epidemic
Mean Response	44.20	41.47	34.99	37.25	35.16
Item Standard Deviation	12.58	14.96	18.90	18.96	20.92
Minimum Item Mean	18.55	13.41	10.45	10.44	10.05
Maximum Item Mean	64.56	75.24	74.44	75.40	76.35

Table 3

Model Fits and Parameter Estimates

for the Final Model

	Experiment				
	1	2	3-Abstract	3-Stock	3-Epidemic
Fit of the Model					
R-squared	.85	.84	.90	.88	.90
RMSD	4.80	6.01	6.04	6.39	6.72
Parameter Estimates					
Joint Presence	9.65	8.22	9.26	9.22	11.30
Violation of Sufficiency	—[a]	-12.97	-8.17	-10.05	-8.42
Violation of Necessity	—[a]	-12.17	-8.07	-5.75	-6.44
Joint Absence	4.03	9.42	2.87	2.52	3.79
Strength of Alternatives	-3.90	-5.69	-3.15	-3.30	-3.20
Regression Constant	29.81	38.46	33.60	36.20	29.96

[a]Not relevant in this experiment.

Table 4

R-Squared Values for

Proposed and Alternative Models

	Experiment				
	1	2	3-Abstract	3-Stock	3-Epidemic
Number of occurrences of each evidence type (Proposed Model)	.85	.84	.90	.88	.90
Count of each evidence type over number of situations	.79	.75	.86	.87	.85
Count of each evidence type over number of informative situations	.79	.72	.76	.75	.73
Count of each evidence type over number of situations with the same outcome	.73	.70	.76	.80	.78
Count of each evidence type over number of informative situations with the same outcome	—	.64	.72	.75	.72
Count of each evidence type over number of possible causes observed in each situation	.71	.81	.82	.81	.81

Technical Reports Presently in this Series

NR 150-412, ONR Contract N0001478C0025

#1. Sternberg, R. J. Intelligence research at the interface between differential and cognitive psychology: Prospects and proposals. January, 1978.

#2. Sternberg, R. J. Isolating the components of intelligence. January, 1978.

#3. Sternberg, R. J., Guyote, M. J., & Turner, M. E. Deductive reasoning. January, 1978.

#4. Sternberg, R. J. Toward a unified componential theory of human reasoning. April, 1978.

#5. Guyote, M. J., & Sternberg, R. J. A transitive-chain theory of syllogistic reasoning. April, 1978.

#6. Sternberg, R. J., & Turner, M. E. Components of syllogistic reasoning. April, 1978.

#7. Sternberg, R. J., Tourangeau, R., & Nigro, G. Metaphor, induction, and social policy: The convergence of macroscopic and microscopic views. April, 1978.

#8. Sternberg, R. J. A proposed resolution of curious conflicts in the literature on linear syllogistic reasoning. June, 1978.

#9. Sternberg, R. J. The nature of mental abilities. June, 1978.

#10. Sternberg, R. J. Psychometrics, mathematical psychology, and cognition: Confessions of a closet psychometrician. June, 1978.

#11. Tourangeau, R., & Sternberg, R. J. Understanding and appreciating metaphors. June, 1978.

#12. Sternberg, R. J. Representation and process in transitive inference. October, 1978.

#13. Tourangeau, R., & Sternberg, R. J. Aptness in metaphor. October, 1978.

#14. Sternberg, R. J. Contrasting conceptions of intelligence and their educational implications. November, 1978.

#15. Sternberg, R. J., & Weil, E. M. An aptitude-strategy interaction in linear syllogistic reasoning. April, 1979.

#16. Sternberg, R. J. Intelligence tests in the year 2000: What forms will they take and what purposes will they serve? April, 1979.

#17. Sternberg, R. J. New views on IQs: A silent revolution of the 70s. April, 1979.

Technical Reports Presently in this Series

NR 150-412, ONR Contract N0001478C0025

#18. Sternberg, R. J., & Gardner, M. K. Unities in inductive reasoning. October, 1979.

#19. Sternberg, R. J. Components of human intelligence. October, 1979.

#20. Sternberg, R. J. The construct validity of aptitude tests: An information-processing assessment. October, 1979.

#21. Schustack, M. W., & Sternberg, R. J. Evaluation of evidence in causal inference. October, 1979.

Navy

1 Dr. Ed Aiken
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 Dr. Robert Breaux
Code N-71
NAVTRAEQUIPCEN
Orlando, FL 32813

1 MR. MAURICE CALLAHAN
Pers 23a
Bureau of Naval Personnel
Washington, DC 20370

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 CDR John Ferguson, MSC, USN
Naval Medical R&D Command (Code 44)
National Naval Medical Center
Bethesda, MD 20014

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 CAPT. D.M. GRAGG, MC, USN
HEAD, SECTION ON MEDICAL EDUCATION
UNIFORMED SERVICES UNIV. OF THE
HEALTH SCIENCES
6917 ARLINGTON ROAD
BETHESDA, MD 20014

1 LT Steven D. Harris, MSC, USN
Code 6021
Naval Air Development Center
Warminster, Pennsylvania 18974

Navy

1 CDR Robert S. Kennedy
Naval Aerospace Medical and
Research Lab
Box 29407
New Orleans, LA 70189

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 CHAIRMAN, LEADERSHIP & LAW DEPT.
DIV. OF PROFESSIONAL DEVELOPMMENT
U.S. NAVAL ACADEMYY
ANNAPOLIS, MD 21402

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 CAPT Richard L. Martin
USS Francis Marion (LPA-Z49)
FPO New York, NY 09501

1 Dr. James McBride
Code 301
Navy Personnel R&D Center
San Diego, CA 92152

2 Dr. James McGrath
Navy Personnel R&D Center
Code 306
San Diego, CA 92152

1 DR. WILLIAM MONTAGUE
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

Navy

1 Commanding Officer
Naval Health Research
Center
Attn: Library
San Diego, CA 92152

1 Naval Medical R&D Command
Code 44
National Naval Medical Center
Bethesda, MD 20014

1 CAPT Paul Nelson, USN
Chief, Medical Service Corps
Code 7
Bureau of Medicine & Surgery
U. S. Department of the Navy
Washington, DC 20372

1 Library
Navy Personnel R&D Center
San Diego, CA 92152

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 JOHN OLSEN
CHIEF OF NAVAL EDUCATION &
TRAINING SUPPORT
PENSACOLA, FL 32509

1 Psychologist
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

Navy

1 Office of Naval Research
Code 441
800 N. Quincy Street
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
OFFICE OF NAVAL RESEARCH BRANCH
223 OLD MARYLEBONE ROAD
LONDON, NW, 15TH ENGLAND

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc
(OP-102)
Washington, DC 20350

1 Scientific Advisor to the Chief of
Naval Personnel (Pers-Or)
Naval Bureau of Personnel
Room 4410, Arlington Annex
Washington, DC 20370

1 LT Frank C. Petho, MSC, USNR (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

Navy

1 Roger W. Remington, Ph.D
Code L52
NAMRL
Pensacola, FL 32508

1 Dr. Bernard Rimland
Navy Personnel R&D Center
San Diego, CA 92152

1 Mr. Arnold Rubenstein
Naval Personnel Support Technology
Naval Material Command (08T244)
Room 1044, Crystal Plaza #5
2221 Jefferson Davis Highway
Arlington, VA 20360

1 Dr. Worth Scanland
Chief of Naval Education and Training
Code N-5
NAS, Pensacola, FL 32508

1 A. A. SJOHOLM
TECH. SUPPORT, CODE 201
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Mr. Robert Smith
Office of Chief of Naval Operations
OP-987E
Washington, DC 20350

1 Dr. Alfred F. Smode
Training Analysis & Evaluation Group
(TAEG)
Dept. of the Navy
Orlando, FL 32813

1 CDR Charles J. Theisen, JR. MSC, USN
Head Human Factors Engineering Div.
Naval Air Development Center
Warminster, PA 18974

1 W. Gary Thomson
Naval Ocean Systems Center
Code 7132
San Diego, CA 92152

Navy

1 Dr. Ronald Weitzman
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

Army

1 Technical Director
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 HQ USAREUE & 7th Army
ODCSOPS
USAAREUE Director of GED
APO New York 09403

1 DR. RALPH DUSEK
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Myron Fischl
U.S. Army Research Institute for the Social and Behavioral Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Ed Johnson
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

1 Dr. Michael Kaplan
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Milton S. Katz
Individual Training & Skill Evaluation Technical Area
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Beatrice J. Farr
Army Research Institute (PERI-OK)
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Milt Maier
U.S. ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

Army

1 Dr. Harold F. O'Neil, Jr.
ATTN: PERI-OK
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

1 Dr. Robert Sasmor
U. S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Frederick Steinheiser
U. S. Army Reserch Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

1 Dr. Joseph Ward
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Air Force

1 Air Force Human Resources Lab
AFHRL/PED
Brooks AFB, TX 78235

1 Air University Library
AUL/LSE 76/443
Maxwell AFB, AL 36112

1 Dr. Philip De Leo
AFHRL/TT
Lowry AFB, CO 80230

1 Dr. Genevieve Haddad
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 CDR. MERCER
CNET LIAISON OFFICER
AFHRL/FLYING TRAINING DIV.
WILLIAMS AFB, AZ 85224

1 Dr. Ross L. Morgan (AFHRL/ASR)
Wright -Patterson AFB
Ohio 45433

1 Dr. Roger Pennell
AFHRL/TT
Lowry AFB, CO 80230

1 Personnel Analysis Division
HQ USAF/DPXXA
Washington, DC 20330

1 Research Branch
AFMPC/DPMYP
Randolph AFB, TX 78148

1 Dr. Malcolm Ree
AFHRL/PED
Brooks AFB, TX 78235

1 Dr. Marty Rockway (AFHRL/TT)
Lowry AFB
Colorado 80230

Air Force

1 Jack A. Thorpe, Capt, USAF
Program Manager
Life Sciences Directorate
AFOSR
Bolling AFB, DC 20332

1 Brian K. Waters, LCOL, USAF
Air University
Maxwell AFB
Montgomery, AL 36112

Marines

1 H. William Greenup
Education Advisor (E031)
Education Center, MCDEC
Quantico, VA 22134

1 DR. A.L. SLAFKOSKY
SCIENTIFIC ADVISOR (CODE RD-1)
HQ, U.S. MARINE CORPS
WASHINGTON, DC 20380

CoastGuard

1 Mr. Richard Lanterman
PSYCHOLOGICAL RESEARCH (G-P-1/62)
U.S. COAST GUARD HQ
WASHINGTON, DC 20590

1 Dr. Thomas Warm
U. S. Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Other DoD

1 Dr. Stephen Andriole
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

12 Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, VA 22314
Attn: TC

1 Dr. Dexter Fletcher
ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BLVD.
ARLINGTON, VA 22209

1 Dr. William Graham
Testing Directorate
MEPCOM
Ft. Sheridan, IL 60037

1 Military Assistant for Training and
Personnel Technology
Office of the Under Secretary of Defense
for Research & Engineering
Room 3D129, The Pentagon
Washington, DC 20301

1 MAJOR Wayne Sellman, USAF
Office of the Assistant Secretary
of Defense (MRA&L)
3B930 The Pentagon
Washington, DC 20301

Civil Govt

1 Dr. Susan Chipman
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. William Gorham, Director
Personnel R&D Center
Office of Personnel Managment
1900 E Street NW
Washington, DC 20415

1 Dr. Joseph I. Lipson
Division of Science Education
Room W-638
National Science Foundation
Washington, DC 20550

1 Dr. Joseph Markowitz
Office of Research and Development
Central Intelligence Agency
Washington, DC 20205

1 Dr. John Mays
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Arthur Melmed
National Intitute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Andrew R. Molnar
Science Education Dev.
and Research
National Science Foundation
Washington, DC 20550

1 Dr. Jeffrey Schiller
National Institute of Education
1200 19th St. NW
Washington, DC 20208

Civil Govt

1 Dr. H. Wallace Sinaiko
Program Director
Manpower Research and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

1 Dr. Thomas G. Sticht
Basic Skills Program
National Institute of Education
1200 19th Street NW
Washington, DC 20208

1 Dr. Frank Withrow
U. S. Office of Education
400 6th Street SW
Washington, DC 20202

1 Dr. Joseph L. Young, Director
Memory & Cognitive Processes
National Science Foundation
Washington, DC 20550

Non Govt

1 Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

1 Dr. John R. Anderson
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. John Annett
Department of Psychology
University of Warwick
Coventry CV4 7AL
ENGLAND

1 DR. MICHAEL ATWOOD
SCIENCE APPLICATIONS INSTITUTE
40 DENVER TECH. CENTER WEST
7935 E. PRENTICE AVENUE
ENGLEWOOD, CO 80110

1 1 psychological research unit
Dept. of Defense (Army Office)
Campbell Park Offices
Canberra ACT 2600, Australia

1 Dr. Alan Baddeley
Medical Research Council
Applied Psychology Unit
15 Chaucer Road
Cambridge CB2 2EF
ENGLAND

1 Dr. Patricia Baggett
Department of Psychology
University of Denver
University Park
Denver, CO 80208

1 Dr. Jackson Beatty
Department of Psychology
University of California
Los Angeles, CA 90024

1 Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

Non Govt

1 Dr. Nicholas A. Bond
Dept. of Psychology
Sacramento State College
600 Jay Street
Sacramento, CA 95819

1 Dr. Lyle Bourne
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. Robert Brennan
American College Testing Programs
P. O. Box 168
Iowa City, IA 52240

1 Dr. John S. Brown
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 DR. C. VICTOR BUNDERSON
WICAT INC.
UNIVERSITY PLAZA, SUITE 10
1160 SO. STATE ST.
OREM, UT 84057

1 Dr. John B. Carroll
Psychometric Lab
Univ. of No. Carolina
Davie Hall 013A
Chapel Hill, NC 27514

1 Dr. William Chase
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1 Dr. Micheline Chi
Learning R & D Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213

1 Dr. John Chiorini
Litton-Mellonics
Box 1286
Springfield, VA 22151

Non Govt

1 Dr. Kenneth E. Clark
College of Arts & Sciences
University of Rochester
River Campus Station
Rochester, NY 14627

1 Dr. Norman Cliff
Dept. of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1 Dr. Allan M. Collins
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Ma 02138

1 Dr. Meredith Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L Street N. W.
Washington, DC 20037

1 Dr. Ruth Day
Center for Advanced Study
in Behavioral Sciences
202 Junipero Serra Blvd.
Stanford, CA 94305

1 Dr. Emmanuel Donchin
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Dr. Hubert Dreyfus
Department of Philosophy
University of California
Berkely, CA 94720

1 Dr. Marvin D. Dunnette
N492 Elliott Hall
Dept. of Psychology
Univ. of Minnesota
Minneapolis, MN 55455

1 ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

Non Govt

1 MAJOR I. N. EVONIC
CANADIAN FORCES PERS. APPLIED RESEARCH
1107 AVENUE ROAD
TORONTO, ONTARIO, CANADA

1 Dr. Ed Feigenbaum
Department of Computer Science
Stanford University
Stanford, CA 94305

1 Dr. Richard L. Ferguson
The American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
Advanced Research Resources Organ.
Suite 900
4330 East West Highway
Washington, DC 20014

1 Dr. John R. Frederiksen
Bolt Beranek & Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Alinda Friedman
Department of Psychology
University of Alberta
Edmonton, Alberta
CANADA T6G 2J9

1 Dr. R. Edward Geiselman
Department of Psychology
University of California
Los Angeles, CA 90024

1 DR. ROBERT GLASER
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

Non Govt

1 Dr. Ira Goldstein
XEROX Palo Alto Research Center
3333 Coyote Road
Palo Alto, CA 94304

1 DR. JAMES G. GREENO
LRDC
UNIVERSITY OF PITTSBURGH
3939 O'HARA STREET
PITTSBURGH, PA 15213

1 Dr. Ron Hambleton
School of Education
University of Massechusetts
Amherst, MA 01002

1 Dr. Harold Hawkins
Department of Psychology
University of Oregon
Eugene OR 97403

1 Dr. Barbara Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. Frederick Hayes-Roth
The Rand Corporation
1700 Main Street
Santa Monica, CA 90406

1 Dr. James R. Hoffman
Department of Psychology
University of Delaware
Newark, DE 19711

1 Dr. Lloyd Humphreys
Department of Psychology
University of Illinois
Champaign, IL 61820

1 Library
HumRRO/Western Division
27857 Berwick Drive
Carmel, CA 93921

Non Govt

1 Dr. Earl Hunt
Dept. of Psychology
University of Washington
Seattle, WA 98105

1 Mr. Gary Irving
Data Sciences Division
Technology Services Corporation
2811 Wilshire Blvd.
Santa Monica CA 90403

1 Dr. Steven W. Keele
Dept. of Psychology
University of Oregon
Eugene, OR 97403

1 Dr. Walter Kintsch
Department of Psychology
University of Colorado
Boulder, CO 80302

1 Dr. David Kieras
Department of Psychology
University of Arizona
Tuscon, AZ 85721

1 Dr. Stephen Kosslyn
Harvard University
Department of Psychology
33 Kirkland Street
Cambridge, MA 02138

1 Mr. Marlin Kroger
1117 Via Goleta
Palos Verdes Estates, CA 90274

1 LCOL. C.R.J. LAFLEUR
PERSONNEL APPLIED RESEARCH
NATIONAL DEFENSE HQS
101 COLONEL BY DRIVE
OTTAWA, CANADA K1A OK2

1 Dr. Jill Larkin
Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

Non Govt

1 Dr. Alan Lesgold
Learning R&D Center
University of Pittsburgh
Pittsburgh, PA 15260

1 Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

1 Dr. Frederick M. Lord
Educational Testing Service
Princeton, NJ 08540

1 Dr. Richard B. Millward
Dept. of Psychology
Hunter Lab.
Brown University
Providence, RI 82912

1 Dr. Allen Munro
Univ. of So. California
Behavioral Technology Labs
3717 South Hope Street
Los Angeles, CA 90007

1 Dr. Donald A Norman
Dept. of Psychology C-009
Univ. of California, San Diego
La Jolla, CA 92093

1 Dr. Melvin R. Novick
Iowa Testing Programs
University of Iowa
Iowa City, IA 52242

1 Dr. Jesse Orlansky
Institute for Defense Analysis
400 Army Navy Drive
Arlington, VA 22202

1 Dr. Robert Pachella
Department of Psychology
Human Performance Center
330 Packard Road
Ann Arbor, MI 48104

Non Govt

1 Dr. Seymour A. Papert
Massachusetts Institute of Technology
Artificial Intelligence Lab
545 Technology Square
Cambridge, MA 02139

1 Dr. James A. Paulson
Portland State University
P.O. Box 751
Portland, OR 97207

1 MR. LUIGI PETRULLO
2431 N. EDGEWOOD STREET
ARLINGTON, VA 22207

1 DR. STEVEN M. PINE
4950 Douglas Avenue
Golden Valley, MN 55416

1 Dr. Martha Polson
Department of Psychology
University of Colorado
Boulder, CO 80302

1 DR. PETER POLSON
DEPT. OF PSYCHOLOGY
UNIVERSITY OF COLORADO
BOULDER, CO 80302

1 DR. DIANE M. RAMSEY-KLEE
R-K RESEARCH & SYSTEM DESIGN
3947 RIDGEMONT DRIVE
MALIBU, CA 90265

1 MIN. RET. M. RAUCH
P II 4
BUNDESMINISTERIUM DER VERTEIDIGUNG
POSTFACH 161
53 BONN 1, GERMANY

1 Dr. Peter B. Read
Social Science Research Council
605 Third Avenue
New York, NY 10016

Non Govt

1 Dr. Mark D. Reckase
Educational Psychology Dept.
University of Missouri-Columbia
12 Hill Hall
Columbia, MO 65201

1 Dr. Fred Reif
SESAME
c/o Physics Department
University of California
Berkely, CA 94720

1 Dr. Andrew M. Rose
American Institutes for Research
1055 Thomas Jefferson St. NW
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Ernst Z. Rothkopf
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

1 Dr. David Rumelhart
Center for Human Information Processing
Univ. of California, San Diego
La Jolla, CA 92093

1 PROF. FUMIKO SAMEJIMA
DEPT. OF PSYCHOLOGY
UNIVERSITY OF TENNESSEE
KNOXVILLE, TN 37916

1 Dr. Irwin Sarason
Department of Psychology
University of Washington
Seattle, WA 98195

1 DR. WALTER SCHNEIDER
DEPT. OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS
CHAMPAIGN, IL 61820

Non Govt

1 Dr. Richard Snow
School of Education
Stanford University
Ctanford, CA 94305

1 DR. ALBERT STEVENS
BOLT BERANEK & NEWMAN, INC.
50 MOULTON STREET
CAMBRIDGE, MA 02138

1 DR. PATRICK SUPPES
INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CA 94305

1 Dr. Hariharan Swaminathan
Laboratory of Psychometric and Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

1 Dr. Brad Sympson
Office of Data Analysis Research
Educational Testing Service
Princeton, NJ 08541

1 Dr. Kikumi Tatsuoka
Computer Based Education Research Laboratory
252 Engineering Research Laboratory
University of Illinois
Urbana, IL 61801

1 Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

1 Dr. John Thomas
IBM Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

1 DR. PERRY THORNDYKE
THE RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CA 90406

Non Govt

1 Dr. J. Uhlaner
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

1 Dr. Benton J. Underwood
Dept. of Psychology
Northwestern University
Evanston, IL 60201

1 Dr. Howard Wainer
Bureau of Social SCience Research
1990 M Street, N. W.
Washington, DC 20036

1 Dr. Phyllis Weaver
Graduate School of Education
Harvard University
200 Larsen Hall, Appian Way
Cambridge, MA 02138

1 Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

1 DR. SUSAN E. WHITELY
PSYCHOLOGY DEPARTMENT
UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66044

1 Dr. J. Arthur Woodward
Department of Psychology
University of California
Los Angeles, CA 90024

1 Dr. Karl Zinn
Center for research on Learning and Teaching
University of Michigan
Ann Arbor, MI 48104